037230-3-T

THE UNIVERSITY OF MICHIGAN



ENGINEERING PSYCHOLOGY LABORATORY

# Ratio Versus Magnitude Estimates of Importance Factors

*Technical Report*

GREGORY W. FISCHER
and
CAMERON R. PETERSON



***Prepared for:***

Engineering Psychology Programs
Office of Naval Research
The Department of the Navy
Arlington, Virginia
Contract No. N00014-67-A-0181-0034
NR 197-014



*Administered through:*

*June 1972*

OFFICE OF RESEARCH ADMINISTRATION · ANN ARBOR

# RATIO VERSUS MAGNITUDE ESTIMATES

# OF IMPORTANCE FACTORS

Technical Report

1 September 1972

Gregory W. Fischer

and

Cameron R. Peterson

Engineering Psychology Laboratory

The University of Michigan

Ann Arbor, Michigan

This research was supported by the Engineering Psychology Programs, Office of Naval Research, under Contract Number N00014-67-A-0181-0034, Work Unit Number NR 197-014.

## TABLE OF CONTENTS

Security Classification

## DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

**1. ORIGINATING ACTIVITY (Corporate author)**

Department of Psychology
University of Michigan
Ann Arbor, Michigan

**2a. REPORT SECURITY CLASSIFICATION**

Unclassified

**2b. GROUP**

**3. REPORT TITLE**

Ratio Versus Magnitude Estimates of Importance Factors

**4. DESCRIPTIVE NOTES (Type of report and inclusive dates)**

Technical Report

**5. AUTHOR(S) (First name, middle initial, last name)**

Gregory W. Fischer
Cameron R. Peterson

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| 1 September 1972 | 10 | 16 |

**8a. CONTRACT OR GRANT NO.**

N00014-67-A-0181-0034

**b. PROJECT NO.**

NR 197-014

**c.**

**d.**

**9a. ORIGINATOR'S REPORT NUMBER(S)**

037230-3-T

**9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)**

None

**10. DISTRIBUTION STATEMENT**

Approved for public release; distribution unlimited.

**11. SUPPLEMENTARY NOTES**

**12. SPONSORING MILITARY ACTIVITY**

Engineering Psychology Programs
Office of Naval Research

**13. ABSTRACT**

Optimal decision making requires that the decision maker trade off various goals or objectives against one another in selecting a course of action. This experiment compared two procedures for assigning importance weights to objectives. The first used magnitude estimates and the second ratio comparisons of importance. The ratio procedure produced substantially greater discrimination between the importance weights assigned to objectives than did the magnitude estimation procedure.

A sensitivity analysis revealed, however, that additive evaluation models were relatively insensitive to the differences between the importance weights produced by the two procedures. Additive models based upon the two types of weights assigned very similar overall values to alternatives.

Security Classification

| 14. KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| Decision Making<br>Trade-offs<br>Decomposed Evaluation<br>Additive Evaluation Models<br>Importance Weights | | | | | | |

# INTRODUCTION

Decision makers must frequently choose among courses of action which lead to the attainment of multiple goals or objectives. Seldom, however, will it be the case that one alternative is best with respect to all objectives. Thus, decision makers must trade-off one objective against another in determining which alternative is most desirable in an overall sense. Until recently this problem of weighing objectives against one another was viewed as inherently subjective in nature and beyond the scope of formal analysis. Recent studies of the subjective evaluation process, however, have revealed two major limitations of the purely subjective approach. First, subjective evaluation is characterized by a substantial degree of random error (Bowman, 1963; Slovic and Lichtenstein, 1971). In addition, decision makers seem unable to take into account more than a few value relevant considerations at a time, thus ignoring potentially important information (Slovic and Lichtenstein, 1971).

Decomposed evaluation procedures have been proposed as a means for improving upon subjective evaluation. The essence of this approach is to divide the evaluation process into a set of simpler subtasks, each of which is well with the judgmental capacities of the decision maker. Given a set of alternatives to be evaluated, decomposition procedures usually involve the following tasks: 1) List the set of objectives or criteria against which alternatives are to be evaluated; 2) Numerically evaluate each alternative

with respect to each objective; 3) Assign relative importance weights to each of the objectives; and 4) Compute the overall value of each alternative, usually with a weighted sum or product. For example, let $0_{1i}$, $0_{2i}$, ..., $0_{ni}$ be scores reflecting the degree to which alternaitve $A_i$ satisfies objectives $0_1$, $0_2$, ..., $0_n$ respectively, and let $w_1$, $w_2$, ..., $w_n$ be the relative importance factors assigned to these objectives. Then, using the weighted sum formulation, the overall value of alternative $A_i$ is given by

$$V(A_i) = w_1 0_{1i} + \ldots + w_n 0_{ni}.$$

Several procedures are available for constructing decomposed evaluation models (Yntema and Torgerson, 1961; Fishburn, 1965; Raiffa, 1969; Hoepfl and Huber, 1970; Edwards, 1971; Keeney, 1971), and a number of validation studies have attested to both the feasibility and desirability of the approach (Yntema and Torgerson, 1961; Eckenrode, 1965; Yntema and Klem, 1965; Lathrop and Peters, 1969; Hoepfl and Huber, 1970; Huber, Daneshgar, and Ford, 1971). But despite the generally favorable results of these validation studies, the assessment of importance weights has proven to be a problem. Decomposed assessments of importance tend to be more uniformly distributed across objectives than are the implicit weights which underlie purely subjective evaluation (Pollack, 1964; Hoepfl and Huber, 1970). In addition, decomposed weights have been shown to be too flatly distributed across objectives when compared with statistically estimated weights in the presence of a known criterion (Lathrop and Peters, 1969). O'Connor (1972) obtained striking evidence of this problem in his development of an index of water quality.

Water pollution experts assigned importance weights to a list of pollution parameters. Fecal coliform contamination, the most important factor and a potentially severe health hazard, initially received only 1.7 times as much weight as color, a factor of relatively minor aesthetic significance. When confronted with this implication of their assessed weights, the water quality experts reassessed their weights substantially, placing relatively more emphasis on the more crucial parameters.

In most of these applications of the decomposition approach, magnitude estimation procedures have been used to assess weights. Typically, the most important objective is arbitrarily assigned an importance of 100. Other objectives are then assigned weights which reflect their importance relative to the first objective. In using this procedure subejcts seem very reluctant to use numbers below 50, thus producing a flat distribution of weights over objectives.

Similar results have been obtained in probability revision experiments in which subjects are asked to modify their opinions about the likelihood of various hypotheses in light of new data. These studies have revealed that subjects tend to avoid assigning extreme probabilities to hypotheses, even in the face of overwhelming evidence (Du Charme, 1969). In addition, however, it was found that subjects made more extreme judgments when assessing odds ratios than when making magnitude estimates of probabilities (Phillips and Edwards, 1965). Extrapolating back to the context of decomposed evaluation, this result suggests that ratio assessments of importance weights should produce substantially less uniform distributions of importance over objectives than do the standard magnitude estimation procedures. The present study was designed to test this hypothesis.

## METHOD

### Subjects

Sixteen University of Michigan undergraduates served as subjects. All were enrolled in an introductory psychology course, and participation in the experiment contributed to the fulfillment of their course requirements.

### Design

Subjects assigned importance weights to six criteria used in the evaluation of the teaching ability of instructors. Those in one treatment condition first assessed weights using a magnitude estimation procedure, then reassessed their weights using a ratio response mode. Subjects in the other condition made the ratio assessments first, then the magnitude estimates.

### Procedure

The six criteria to which weights were assigned are listed below:

1) Class Discussion: Does the instructor encourage students to ask questions and express their opinions?
2) Fairness: Does the instructor deal with students in a fair and impartial manner?
3) Knowledge: Is the instructor well informed about the subject matter of the course?
4) Organization: Are the instructor's class presentations well prepared and organized?
5) Relevance: Does the instructor relate the course materials to the real life experiences of the student?
6) Responsiveness: Is the instructor responsive to the students needs, feelings and problems?

After becoming familiar with this list of criteria, subjects were randomly divided into two groups of eight each. For the magnitude estimation response mode, subjects first ranked the six criteria in order of importance. They then arbitrarily assigned an importance of 100 to the most important criterion. Next, they assigned relative importance factors to each of the other criteria by making a slash through a ten inch line divided into 100 equal intervals and numbered from 0 to 100. Subjects were instructed to think of these numbers as percentages. For example, a criterion assigned a value of 50 should be 50% as important as the most important criterion.

For the ratio estimates, subjects again ranked the same six criteria in order of importance. They then successively compared the most important criterion with each of the other five by making a slash through a logarithmically spaced ratio scale that ranged from 1:1 to 100:1. Here subjects were instructed that a ratio of 2:1 indicated that the most important criterion was twice as important as the one with which it was being compared. Use of 100:1 as the upper bound of the ratio scale eliminated the possibility that ratio judgments would be more extreme simply because they had no upper bound.

## RESULTS AND CONCLUSIONS

For purposes of data analysis, each of the two sets of weights was normalized to sum to 100. Then, within each subject, the variance of these two sets of weights was computed. These variances are displayed in Table 1. The greater the variance of a set of weights, the less uniform the distribution of importance across attributes. Thus, the hypothesis that ratio assessments of weights result in less uniform distributions of importance across criteria implies that the variance of the ratio assessments should be greater than that of the magnitude assessments. This hypothesis was confirmed at an ordinal level for 15 of 16 subjects. The median variance for the ratio assessments (Vr) was 130.7 whereas the median variance for the magnitude estimates (Vm) was only 23.1. Finally, the median ratio of these variances, Vr/Vm, was 4.5:1.

These results clearly demonstrate that the ratio response mode generates less uniform weight assessments than does the magnitude estimation response mode. The next analysis was designed to determine the importance of this difference in terms of its effect on decisions. A number of previous studies have suggested that additive evaluation models are relatively insensitive to minor variations in weighting parameters; that is, that evaluation models based on different sets of weights will assign very similar overall values to multi-dimensional alternatives (O'Connor, 1972). To test this hypothesis in the present context, a simple numerical analysis was conducted. Because actual evaluations with respect to each of the six criteria were unavailable, a set of scores on the six criteria was randomly generated representing 500

hypothetical instructors. Each criterion was assigned scores between 0 and 100, with scores being randomly generated from a uniform distribution over the range of 0 to 100. The data generating process used for this analysis was such that the criteria were uncorrelated. Next, for each subject, overall scores were assigned to each hypothetical instructor according to the following two models:

$$M_r = r_1X_{1i} + r_2X_{2i} + \dots + r_nX_{ni}$$

$$M_m = m_1X_{1i} + m_2X_{2i} + \dots + m_nX_{ni}.$$

Here, $X_{ji}$ represents the score of the i-th instructor with respect to the j-th criteria, $r_j$ the weight assessed for the j-th criterion using the ratio response mode, and $m_j$ the weight assessed for the j-th criterion using the magnitude estimate response mode. Thus, $M_r$ is the additive evaluation model based upon the ratio assessments and $M_m$ the additive model based upon the magnitude estimates.

Correlations between these two models were computed in order to determine the practical significance of the discrepancies between the two sets of weights. These correlations (presented in Table 1) demonstrate that, except in the case of very severe discrepancies, the models based upon the ratio and magnitude estimates are nearly equivalent. Only four of the 16 correlations are below .90, and in each of these cases the Vr/Vm ratios are extremely high. Over all subjects, the median correlation between the two models is .92. These results indicate that although the two assessment procedures do produce systematically different importance distributions across objectives, the robustness of additive models is so great that these differences will frequently be inconsequential from a practical standpoint.

TABLE 1

Variance of the Two Sets of Weights and Correlations Between Additive Models Based on These Weights

| Subject | Group | Vr | Vm | Vr/Vm | R |
|---|---|---|---|---|---|
| 1 | 1 | 106.1 | 31.6 | 3.3 | .94 |
| 2 | 1 | 12.0 | 7.2 | 1.7 | .99 |
| 3 | 1 | 73.1 | 80.1 | .9 | .99 |
| 4 | 1 | 256.9 | 19.9 | 12.9 | .83 |
| 5 | 1 | 209.8 | 52.4 | 4.0 | .90 |
| 6 | 1 | 53.4 | 23.6 | 2.3 | .98 |
| 7 | 1 | 144.7 | 18.7 | 7.6 | .89 |
| 8 | 1 | 327.0 | 24.3 | 13.5 | .83 |
| 9 | 2 | 175.2 | 34.3 | 5.1 | .92 |
| 10 | 2 | 47.3 | 12.4 | 3.9 | .98 |
| 11 | 2 | 116.7 | 6.1 | 19.1 | .90 |
| 12 | 2 | 207.9 | 88.3 | 2.3 | .92 |
| 13 | 2 | 451.3 | 24.6 | 18.4 | .78 |
| 14 | 2 | 17.6 | 13.0 | 1.3 | 1.00 |
| 15 | 2 | 20.8 | .1 | 208.0 | .97 |
| 16 | 2 | 161.0 | 22.7 | 7.1 | .92 |
| Median | | 130.7 | 23.1 | 4.5 | .92 |

Note: Vr and Vm refer to the variance of the ratio and magnitude estimates respectively. R refers to the correlation between additive models based upon these two sets of weights.

## REFERENCES


Bowman, E. H. Consistency and optimality in managerial decision making. Management Science, 1963, 9, 310-321.

Du Charme, W. M. A response bias explanation of conservative human inference. Doctoral dissertation, The University of Michigan, 1969.

Eckenrode, R. T. Weighting multiple criteria. Management Science, 1965, 12, #3.

Edwards, W. Social utilities. In Decision and Risk Analysis: Powerful New Tools for Management, Proceedings of the Sixth Triennial Symposium, June 1971, Hoboken: The Engineering Economist, 1972, 119-129.

Fishburn, P. C. Independence in utility theory with whole product sets. Operations Research, 1965, 13, 28-45.

Hoepfl, R. T., and Huber, G. P. A study of self-explicated utility models. Behavioral Science, 1970, 15, 408-414.

Huber, G. P., Daneshgar, R., and Ford, D. L. An empirical comparison of five utility models for predicting job preferences. Organizational Behavior and Human Performance, 1971, 6, 267-282.

Keeney, R. L. Utility independence and preference for multi-attributed consequences. Operations Research, 1971, 19, 875-893.

Lathrop, R. G., and Peters, B. E. Subjective cue weighting and decisions in a familiar task. Proceedings, 77th Annual Convention, American Psychological Association, 1969.

O'Connor, M. F. The application of multi-attribute scaling techniques to the development of indices of water quality. Doctoral dissertation, The University of Michigan, 1972.

Phillips, L., and Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-354.

Pollack, I. Action selection and the Yntema-Torgerson worth function. In Information System Science and Engineering: Proceedings of the First Congress of the Information Systems Sciences, New York: McGraw-Hill, 1964.

Raiffa, H. Preferences for multi-attribute alternatives. The Rand Corporation, RM-5868-DOT/RC, April 1969.

Slovic, P., and Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Yntema, D. B., and Klem, L. Telling a computer how to evaluate multi-dimensional situations. IEEE Transactions on Human Factors in Electronics, 1965, HFE-6, 3-13.

Yntema, D. B., and Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions on Human Factors in Electronics, 1961, HFE-2, 20-26.